8th TRB Conference on the Application of Transportation Planning Methods
22-26 April 2001, Corpus Christi, Texas
Session 10: Corridor/Subarea Planning and Simulation

Recommended Guidelines for the Calibration and Validation of Traffic Simulation Models

Ronald T. Milam, AICP & Fred Choa, P.E., *Fehr & Peers Associates, Inc.*

Abstract. Traffic engineers and transportation planners are using traffic simulation models with greater frequency to plan and design future transportation facilities. However, the transportation profession has not established formal and consistent guidelines regarding the development and application of these models. The lack of such guidance or direction has lead to conflicts between model users, inappropriate use of the models, and inaccurate results from the models. Many of these situations could be avoided if guidelines were available that address both model development and application.

The purpose of this paper is to present an initial set of recommended guidelines for the development and application of traffic simulation models. The guidelines are based on previously published information, interviews with practitioners, and results from successfully completed simulation projects. Key issues that will be addressed include the following:

- Calibration of model parameters for traffic control operation, traffic flow characteristics, and driver behavior;
- Validation guidelines for traffic flow measurement; and
- Multiple run requirements for simulation models.

The guidelines contained in the paper require formal refinement, but have been proven successful in the field and were accepted by technical professionals as well as decision makers. Ideally, this paper will be used a building block for more formal guidelines.

The use of traffic simulation models in traffic operations and traffic impact studies is increasing as technology provides traffic engineers and transportation planners with more user-friendly programs and a wider variety of programs to choose. Traffic simulation programs provide an advantage over traditional analysis tools such as the Highway Capacity Software (HCS)(1) in that they can provide an analysis of entire roadway systems and a visual simulation of the results. However, to determine if the results and the visual simulation are accurate requires a high degree of confidence that the model has been calibrated within acceptable tolerances and achieved an acceptable level of validation for key traffic flow characteristics. Unfortunately, the traffic engineering profession has not kept up with the development of software programs when it comes to guidance on the definitions of "calibrated within acceptable tolerances" and "achieved an acceptable level of validation." This paper attempts to provide some initial guidance in this regard with the intent of generating additional discussion and research into the topic.

The remainder of the paper is organized as follows:

- Description of Traffic Simulation Models;
- Calibration of Traffic Control Operations, Traffic Flow Characteristics, and Driver Behavior in Model Development;
- Validation Standards for Traffic Flow Measurement;
- Multiple Run Requirements for Simulation Models; and

- Conclusions.

Description Of Traffic Simulation Models

Traffic simulation models are used in many cases to visually display analysis results. Other common reasons for their use are due to complex or unique roadway geometries that other analysis programs cannot evaluate or a desire to evaluate a combined system of arterial and freeway facilities. One of the better-known traffic simulation programs in the United States is CORSIM(2). This paper focuses on the CORSIM program, which was developed under the guidance of the Federal Highway Administration (FHWA) and is used widely due to its relatively low cost, ability to simulate combined systems of arterial and freeway facilities, user-friendly interface, high-quality visual simulation, and acceptance by reviewing agencies at all levels of U.S. government. Nevertheless, these guidelines are also applicable for other simulation programs. The acceptance issue is important because each simulation model is built upon a basic set of stochastic algorithms that attempt to represent vehicular traffic flow through various types of roadway systems under various conditions. This paper does not evaluate the different algorithms, but points out that reviewing agencies such as the California Department of Transportation (Caltrans) have been more likely to accept the results from a simulation model that was built upon traffic flow conditions measured and modeled in the United States and approved by FHWA.

Because of the stochastic nature of simulation programs, their use requires three basic steps to be completed prior to finalizing any analysis results. First, traffic control operations, traffic flow characteristics, and driver behavior components of the model need to be calibrated to conditions observed or measured in the field. Second, the calibration needs to be linked directly to validation of the model involving a comparison of simulated and observed traffic flow conditions of the system under study. This comparison is intended to provide a direct measurement of how well the model results match existing and observed traffic flow conditions. Third, a statistical analysis of the results is required based on the average values and range of variation generated by multiple runs of the model using varying random number seeds. This effort is necessary due to the stochastic nature of simulation models. Given these three steps are necessary and limited information has been produced to date by model developers or others in the traffic engineering profession, this paper provides some initial guidance for each step based on past simulation model application studies that received approval from various reviewing agencies.

Calibration Of Traffic Simulation Models

The unique aspect of simulation models for traffic operations is that they include a set of independent variables used to describe traffic control operations and driver behavior that are not available in standard traffic operations analysis packages such as HCS. In addition, simulation models include a greater range of adjustment parameters for traffic control devices and traffic flow characteristics compared to other analysis packages. With the introduction of additional variables into the analysis, higher levels of effort and expertise are required to ensure that the simulation model is accurately reflecting observed field conditions. For example, HCS does not contain any input variables related to driver behavior. In CORSIM, the user has the option to adjust a number of variables related to driver behavior including, but not limited to, the following:

- Car-Following Sensitivity Factor (Card Type 20 and 68);
- Lane Change Parameters (Card Type 70 and 81);
- Gaps for Permissive Left-Turns and for Right-Turns on Red or at Signs (Card Type 145); and
- Driver's Familiarity with Paths Distribution (Card Type 153).

This is not a comprehensive list but does provide an indication of the level of detail included in the CORSIM model and how different a simulation model is compared to the traditional programs such as HCS. With a basic understanding of this key difference, it should be clear that CORSIM requires more effort and expertise to ensure that the model replicates existing conditions. The authors recognize that this step is sometimes ignored in traffic operations or traffic impact studies using HCS or other traditional analysis tools. Not only is this unacceptable when using traditional analysis tools, it creates the potential for major errors in the application of simulation models as discussed below.

Recommended Calibration Steps

Calibration is the process by which the individual components of the simulation model are adjusted or tuned so that the model will accurately represent field measured or observed traffic conditions. The unique components or parameters of a simulation model that require calibration include the following:

- Traffic control operations;
- Traffic flow characteristics; and
- Driver behavior.

This step should not be confused with validation; whereby, the accuracy of the model is tested by comparing traffic flow data generated by the model with that collected in the field. Validation is discussed later in the paper and is directly related to the calibration process because adjustments in calibration are often necessary to improve the model's ability to replicate field measured traffic flow conditions.

With regards to calibration, CORSIM and other simulation models contain numerous variables to describe traffic control operation, traffic flow characteristics, and driver behavior. Simulation models contain default values for each variable, but also allow a range of user supplied values for each variable. In some cases, the variables affect the entire network while others are specific to individual roadway segments or nodes. Changes to these variables during calibration should be based on field measured or observed conditions. In other words, a change in the variables should be justified and defensible.

Unfortunately, many of the variables used in simulation models are difficult to measure in the field, but can have substantial effect on the model's performance. For example, key driver behavior variables included with the CORSIM program that can substantially influence traffic flow volumes, queues, and speeds are summarized in Table 1. For each variable, the table includes the default value used by CORSIM as well as the effective range of the variable.

The first four parameters in Table 1 are link specific and should be calibrated to observed traffic conditions. Similar to the methodology defined in the Highway Capacity Manual (HCM) (3), the effects of start-up lost time, queue discharge rate, and pedestrian demand have a direct impact on

Table 1. Calibration parameters

Parameters	Effect	Default Value	Calibration Range
Start-up Lost Time (Arterials)	Link Specific	2.0 Seconds	0.5 to 9.9 Seconds
Start-up Lost Time (Freeways)	Link Specific	1.0 Second	0.5 to 6.0 Seconds
Queue Discharge Rate (Arterials)	Link Specific	1.8 Seconds (2,000 vphpl)	1.4 to 2.4 Seconds (1,500 – 2,270 vphpl)
Pedestrian Demand	Link Specific	0 (If left blank) No pedestrians	Demand 0 to 4 0 – 500 pedestrians
Car-Following Sensitivity Factor (Freeways)	Network-wide	Driver Type 1 (.6 Seconds) to Driver Type 10 (1.5 Seconds)	0.6 to 1.5 Seconds for Driver Types 1 to 10
Time to complete a lane change (Freeways)	Network-wide	3.0 Seconds	2.0 to 5.0 Seconds
Acceptable gap in oncoming traffic for permissive left-turners (Arterials)	Network-wide	Driver Type 1 (7.8 Seconds) to Driver Type 10 (2.7 Seconds)	2.7 to 7.8 Seconds for Driver Types 1 to 10
Acceptable gap in oncoming traffic for right-turns-on-red or right-turn at stop sign (Arterials)	Network-wide	Driver Type 1 (10.0 Seconds) to Driver Type 10 (3.6 Seconds)	3.6 to 10.0 Seconds for Driver Types 1 to 10
Driver's Familiarity with Network (Arterials)	Network-wide	10 Percent Recreational 90 Percent Commuter	Sum of recreational and commuter must equal 100

the operating conditions of a signalized intersection. In addition, these factors can vary on an intersection-by-intersection basis, and relying on the default values can result in erroneous delays, queue lengths, and level of service (LOS) estimates. Therefore, calibrating the link specific information is critical in validating existing conditions and accurately forecasting future operating conditions. For freeway conditions, the link specific start-up lost time is also important in calibrating operating conditions of the freeway under stop-and-go conditions.

The next five parameters are network-wide calibration values. For the freeway network, the default car-following sensitivity factors (0.60 to 1.50 seconds) can be adjusted to reflect the actual driving behavior of the traffic stream. For the arterial network, acceptable gaps for left- and right-turning vehicles should also be calibrated based on field observations and the degree of driver's acceptance of gaps in the traffic stream. Driver familiarity with area is another calibration factor that can affect the operations of the study corridor and the benefits of signal coordination and geometric modifications.

By using the calibration factors identified in Table 1, the accuracy of the existing conditions CORSIM network should be significantly improved and replicate "real world" traffic conditions. If the calibration process is completed accurately, viewing the traffic simulation should produce "What You See is What You Should Get (WYSIWYSG)" results.

Validation Standards For Traffic Simulation Models

Under ideal conditions, the calibration of individual components of a simulation model will improve the simulation model's ability to replicate traffic flow results that match field conditions within an acceptable range of error. Typical traffic flow characteristics that can be used in validation include traffic volumes, average travel time, average travel speed, queue lengths, and density. Unfortunately, professional guidelines that define the acceptable range of error for these characteristics have not been developed. Instead, transportation professionals have either ignored the need for validation or developed their own guidelines. Examples of validation guidelines used in recent projects by the authors and accepted by agencies such as Caltrans are contained in Table 2. Although these guidelines are a starting point for discussing guidelines for the transportation profession, they lack statistical justification to determine if they provide an acceptable range of error.

To understand how the validation guidelines in Table 2 can be applied, the following example was prepared using the CORSIM Automated Results System (CARS) post-processor developed by Fehr & Peers Associates, Inc. It should be noted that a critical component of the CARS post-processor is that it converts CORSIM output into performance measures that are consistent with the analysis methodology contained in the 1997 Highway Capacity Manual (HCM)(3). The need for this conversion is the subject of another paper by the authors for the 2000 ITE District 6 Conference entitled, "HCM & CORSIM - Resolving the Differences." Further, any analysis of CORSIM output requires that average results and the variation in the results are calculated and reported. This requirement is discussed in detail in the next section and is a built-in function of the CARS post-processor.

Exhibit 1 presents the results from the CARS post-processor for a signalized intersection. The CARS program provides standard HCM level of service (LOS) results, with the individual lane group delay, approach delay, and overall intersection delay for multiple CORSIM runs. The program also determines the percent of demand served (i.e., the input volume divided by the CORSIM output volume) for each movement and the entire intersection. The first step of the validation process is to determine that CORSIM assigned an acceptable portion of the input volume. Because turning movements in CORSIM are based on percentages and not absolute volumes, input and output volumes will not be identical in all cases. However, the guidelines in Table 1 suggest that the output volumes should be within 95 to 105 percent of the input volumes. This guideline is met for the overall intersection in the example below, but some individual turning movements fall outside this range, which indicates that additional model refinement may be necessary.

Table 2. Validation guidelines

Parameters	Description	Validation Criteria
Volume served	Percent difference between input volume and CORSIM output or assigned volume	95 to 105% of observed value
Average Travel Time	Standard deviation between floating car average travel times and CORSIM simulated travel time for a series of links	1 standard deviation
Average Travel Speed	Standard deviation between floating car average speed and CORSIM simulated average travel speed for individual links	1 standard deviation
Freeway Density	Percent difference between observed freeway density (from volume counts and floating car travel speed) and CORSIM simulated density	90 to 110% of observed value
Average and Maximum Vehicle Queue Length	Percent difference between observed queue lengths and CORSIM simulated queue lengths	80 to 120% of observed value

Operations Summary

	Volume Demand	Volume Served	%	Delay Time (veh mins)	Lane Group (sec/veh)	Approach Delay/LOS	
NB Left	110	97	88	68	42		
NB Thru	2274	2271	99	752	19	Northbound	
NB Right	41	57	139	20	21	20.9	C
SB Left	28	23	82	10	27		
SB Thru	2022	2041	100	464	13	Southbound	
SB Right	67	55	82	1	1	13.4	B
EB Left	220	217	98	100	27		
EB Thru	803	805	100	481	35	Eastbound	
EB Right	0	0	0	0	0	34.1	D
WB Left	330	337	102	223	39		
WB Thru	0	0	0	0	0	Westbound	
WB Right	34	26	76	5	12	37.3	D
TOTAL	5929	5929	100	2128	127687	21.5	C

Exhibit 1. CARS Level of Service / Volume Served Results

The second step of the validation process is to compare the observed average travel time and travel speed with CORSIM simulated results. Using the car following technique, a minimum of 5 to 10 travel time runs through the project study area should be conducted as part of the data collection effort. Using the observed travel times, the average travel time and travel speed can be determined. In addition, the standard deviation can also be determined. Exhibit 2 presents the results of the average travel time and travel speed for multiple CORSIM runs. In this example, the average time to travel approximately ¼ mile on westbound Mt. Diablo Boulevard (through the downtown area of the Town of Lafayette, CA) is 1 minute 34 seconds, at an average travel speed of 11.0 miles per hour. The exhibit also presents the results of three additional travel paths through the project study area. Using the guidelines in Table 1, the model results would be acceptable if the simulated average travel time and travel speed results are within 1 standard deviation of the field collected data.

For the freeway network, the results of the volume count and time/speed data can be used to determine the existing density of the freeway segment. The field observed density is then compared to the CORSIM simulated density. Based on the guidelines in Table 1, the model results would be acceptable if the simulated density is within 10 percent of the observed freeway density.

The last, and most difficult, validation criteria are the average and maximum queue length. Based on the definition for average queue length, CORSIM determines the average queue length on a roadway segment based on 5-second histograms. CORSIM stores the vehicle content in each lane every 5 seconds during the simulation and calculates the average over the entire peak hour. For the maximum queue length, the highest vehicle content at any point during the simulation is recorded. During the data collection process, the maximum queue length is recorded at the beginning of the each green phase during the peak hour. From the observed data, the average queue length at the beginning of the green phase can be calculated.

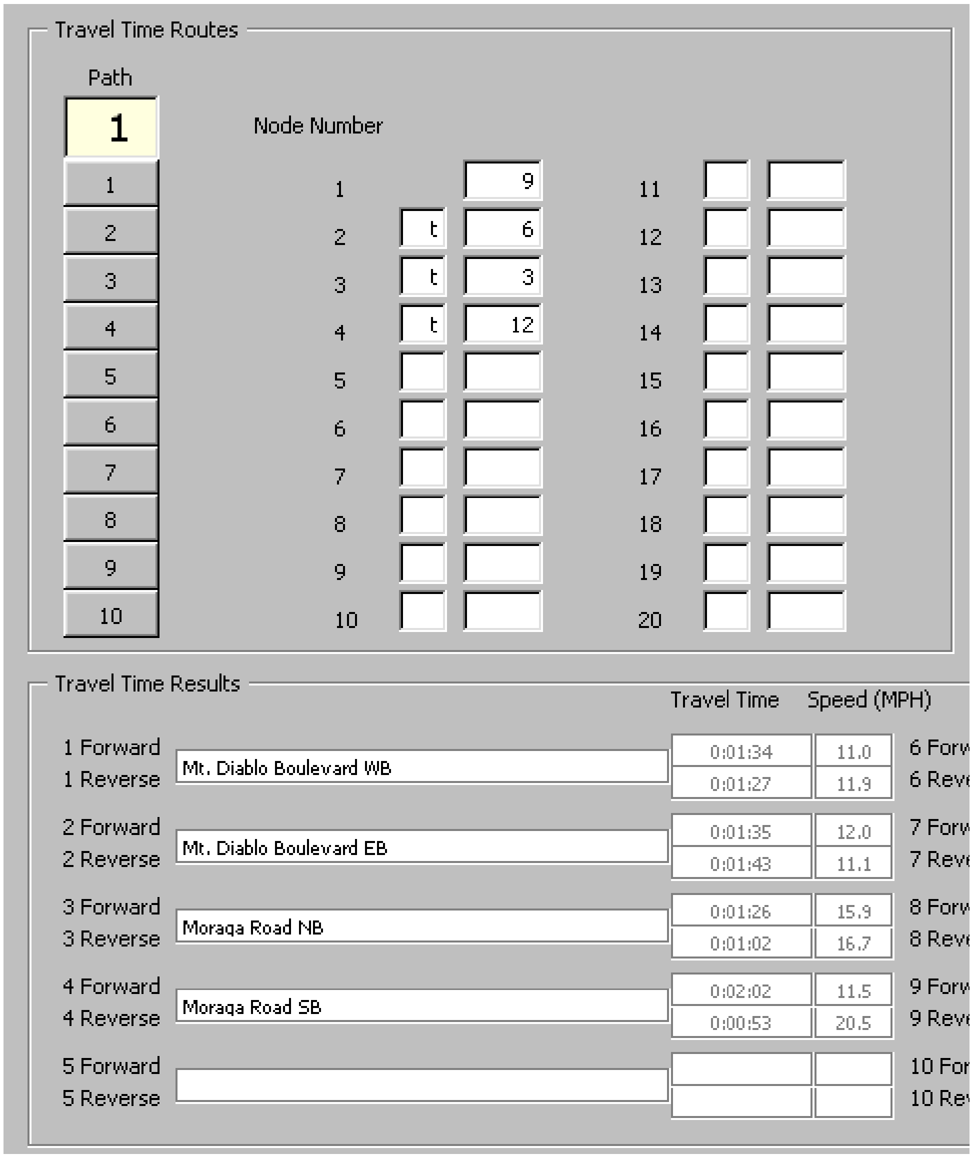

Exhibit 2. CARS Average Travel Time / Travel Speed Results

Exhibit 3 presents the results of the CORSIM simulated maximum queue. As stated above, the maximum queue represents the highest vehicle content during any 5-second interval in the CORSIM simulation. Theoretically, this should occur at the beginning of the green phase. But additional factors (including driveway operations, incidents, and parking maneuvers) can cause the maximum queue length to occur at other times. Therefore, validation of the average and maximum queue length is based on a combination of the validation criteria and viewing the simulation. At a minimum, the CORSIM model should match what you see in the field and the maximum queue length should be within 20 percent of the CORSIM simulated vehicle queue length.

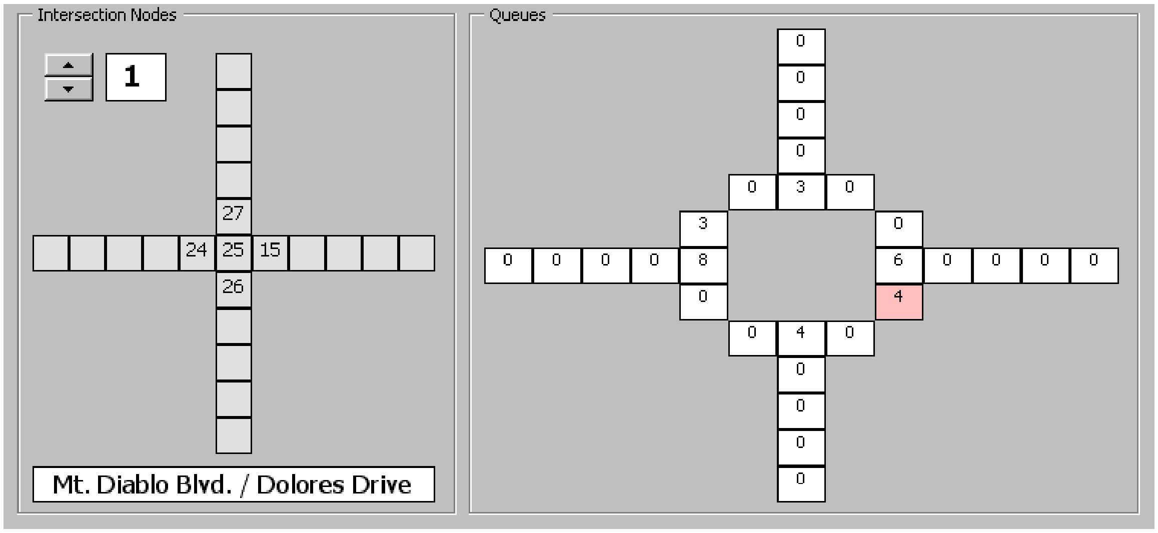

Exhibit 3. CARS Maximum Queue (Vehicles) Results

The validation process is integral to the analysis of existing and future traffic operations and in the consensus building process. If the CORSIM results do not accurately represent "real-world traffic conditions" and the experiences of concerned stakeholders (residents, business owners, and public officials), then the credibility of recommendations derived from the simulation model is lost.

Multiple Run Requirements For Simulation Models

CORSIM and other traffic simulation programs are based on stochastic algorithms (i.e., random processes) that describe driver behavior and traffic operations. Typically, the programs rely on a unique random number seed to generate output results for each individual model execution. With this or any type of statistical model, the user needs to report the average result as well as the amount of variation in the result. A common mistake made by simulation model users in applications is that they only perform one run and then report the results. According to an article by David Hale of McTrans, "This approach equates to throwing a die once, getting a 6, and concluding that each face of the die is a 6."(4) Hale recommends that an estimated average value and confidence interval should be provided for each result under evaluation. One potential method for accomplishing this task is outlined in Hale's article based on a straightforward technique that is applied at the University of Florida. The approach is repeated below for the benefit of potential users:

> The following technique is used to determine the approximate number of required simulation runs with varying random number seeds. After having performed the appropriate number of runs, the user may estimate that the true expected result lies inside the computed interval with 90 percent confidence.
>
> Example: Ten initial simulation runs with different random number seeds are performed and the output results are listed in the following table. Between 10 and 20 initial runs are recommended, and the analysis period should be as long as possible, although 60-minute periods are often used in the academic environment. (Ten long runs are preferable to twenty short runs.)

Replication	Total Delay In Queue (mins)	(avg-obs)	$(\text{avg-obs})^2$
1	1.53	0.50	0.25
2	1.66	0.37	0.14
3	1.24	0.79	0.62
4	2.34	-0.31	0.10
5	2.0	0.03	0.00
6	1.69	0.34	0.12
7	2.69	-0.66	0.44
8	2.86	-0.83	0.69
9	1.70	0.33	0.11
10	2.60	-0.57	0.32
	Avg = 2.03		Sum = 2.79

From the table above, the calculated average value of 2.03 minutes of total delay in queue, and the summation of the $(\text{average-observation})^2$ terms, 2.79, are used in the upcoming calculations. The objective is to obtain a point estimate and approximate 90 percent confidence interval for the expected total delay in queue.

First, calculate the sample variance, S^2, for n=10 runs:

$$S^2(10) = 2.79/(n-1) = 2.79/9 = 0.31$$

The 't' factor for 10 runs and 90 percent confidence, along with the sample average and variance, are now used to obtain the confidence interval estimate. For t9,0.95 = 1.833, the 10-run, 90% confidence interval is:

$$\overline{X}(10) = 2.03, S^2(10) = 0.31$$

$$\overline{X}(10) \pm t_{9,0.95}\sqrt{S^2(10)/10} = 2.03 \pm 0.32$$

After the initial 10 runs, this variation of plus or minus 0.32 appears to be unacceptable when compared to the average of 2.03 minutes in queue. This is because it represents plus or minus 16%, or a confidence interval equal to 32% of the average. If the user had performed only the initial replication, the result could have been anywhere within the range 1.71 to 2.35. Note that while we are 90% confident that the true average lies within that region, only two of the observed results lie within that bound. This points to the need to be cautious in drawing conclusions from a few observations, maybe only one. For the network above, suppose that a confidence interval less than or equal to 10% of the average is desired. Determine the total number of simulation runs needed. For this case, the number of initial runs n1 = 10, variation h1 = 0.32, sample average:

$$\overline{X}(10) = 2.03$$

and confidence factor . The total number of runs (n2) is:

$$n_2 = (10)\left(\frac{0.32}{0.1x2.03}\right)^2 = 25$$

Therefore, an additional 15 simulation runs are needed to get the desired accuracy of the confidence interval. The previous table containing data for replications 1-10 could be reconstructed with the new data upon completion of the additional 15 runs to verify the new 10 percent confidence interval, if desired.

This type of approach provides a high degree of confidence in the simulation results. However, many practitioners may find this level of effort onerous. With CORSIM, the level of effort can be reduced substantially by using the multiple run batch processing that is available as part of the program and then creating a post-processor such as CARS that automates the statistical calculations above. This type of processor should be included with the program or, at a minimum, the information above should be included as part of the user's guide.

Conclusions

Based on the information presented in this paper, the following guidelines are recommended for the application of traffic simulation models:

1. Conduct field measurements and observations of existing traffic flow characteristics and driver behavior for the roadway facilities under study;
2. Calibrate the traffic simulation model to match measured or observed conditions;
3. Validate the traffic simulation model in an iterative process tied to the calibration effort and use guidelines that reduce the risk of under- or over-prediction of model results until professional guidelines are developed; and
4. Average the simulation model results after conducting multiple runs and determine the variation in the results to ensure an adequate number of runs has been completed for the desired level of accuracy.

With this basic approach, the results from the traffic simulation model will be viewed with a high degree of confidence. Establishing this level of confidence is an integral part of establishing consensus about recommendations derived from the simulation model results. To this extent, the transportation profession needs to establish guidelines for the definitions of "calibrated within acceptable tolerances" and "achieved an acceptable level of validation."

References

1. University of Florida, Highway Capacity Software Version 3.1, Gainesville, Florida, 1994-1998.
2. Federal Highway Administration, Traffic Software Integrated System User's Guide Version 4.3, McLean, Virginia, May 1999.
3. Transportation Research Board, Highway Capacity Manual (HCM), Special Report 209, Third Edition, 1997.
4. University of Florida, How Many NETSIM Runs are Enough?, McTrans Newsletter Volume 11 Number 3 March 1997, Gainsville, Florida.